

# Hiddenによるデータの可視化



## 調査の目的

都道府県における健康意識や肥満度にまつわるデータと、  
その他周辺の社会データに関する相関関係を見つける。

# データについて①

以下に、使用したカラムについての説明を載せる。

BMI:(男性(20~69歳)のBMIの平均値+女性(40~69歳)のBMIの平均値)/2 (kg/m<sup>2</sup>)

野菜:(男性(20歳以上)の野菜摂取量の平均値+女性(20歳以上)の野菜摂取量の平均値)/2 (g/日)

食塩:(男性(20歳以上)の食塩摂取量の平均値+女性(20歳以上)の食塩摂取量の平均値)/2 (g/日)

歩数:(男性(20~64歳)の歩数の平均値+女性(20~64歳)の歩数の平均値)/2 (歩/日)

喫煙:男性(20歳以上)の習慣的に喫煙している人の割合 (%)

収入:一人当たり県民所得 (千円)

ラーメン:人口10万人あたりのラーメン店舗数 (店舗/万人)

賃金:最低賃金 (円)

密度:人口密度 (人/km<sup>2</sup>)

車:人口一人当たりの車保有数 (台)

診断:特定健康診断受診率

生活習慣病:生活習慣病患者数/県内の人口

気温:平均気温 (°C)

睡眠:平均睡眠時間 (分)

メタボ:メタボリックシンドローム患者数割合 (%)

クラス:人口密度を3分割にし、下位1/3(180.25未満)に1、真ん中(180.25以上317.65未満)に2、上位1/3(317.65以上)に3を割り当てた

クラスのみCategoryデータで、それ以外はNumericである。

# データについて②

- BMI、野菜、食塩、歩数、喫煙

平成 28 年 国民健康・栄養調査結果の概要 ([https://www.mhlw.go.jp/file/04-Houdouhappyou-10904750-Kenkoukyoku-Gantaisakukenkouzoushinka/kekkgaiyou\\_7.pdf](https://www.mhlw.go.jp/file/04-Houdouhappyou-10904750-Kenkoukyoku-Gantaisakukenkouzoushinka/kekkgaiyou_7.pdf))

- 収入

県民経済計算（平成18年度 - 平成30年度）（2008SNA、平成23年基準計数） 一人当たり県民所得(Excelファイル)

- 賃金

厚生労働省による平成28年度地域別最低賃金時間額答申状況(<https://www.mhlw.go.jp/file/04-Houdouhappyou-11201250-Roudoukijunkyoku-Roudoujoukenseisakuka/280823chiikibetsusaiteichingin-toushin.pdf>)

- 密度、車、生活習慣病、気温、睡眠、人口密度、クラス

e-Stat 政府統計の総合窓口(<https://www.e-stat.go.jp/regional-statistics/ssdsview/prefectures>)

- ラーメン

地域の入れ物、ラーメン店数の都道府県ランキング(<https://region-case.com/rank-office-ramen/>)

- メタボ、診断

厚生労働省 特定健康診査・特定保健指導に関するデータ(特定健康診査、メタボリックシンドローム)

より引用し、使用しやすいデータになるよう一部加工した。

## データについて②

様々な統計データからデータを引用したが、その際に年度を揃えないと相関関係が正しく測れないと思ったため、どのデータも平成28年度のデータを使用することにした。

平成28年度のデータが揃っていない場合は、誤差1年なら急激に変化することはないだろうと思い誤差1年以内のデータを使用することにした。また、熊本県は地震の影響で平成28年度のデータが多くなかったため全体から抜いた。

## データについて③

男女別データの扱いについて:

BMI、野菜摂取量、食塩摂取量、歩数は男女別データ、喫煙率に関しては男性のデータしか存在しなかった。

はじめ、男性のみのデータを用いていたがあまり相関関係が見られなかったため、男女別データがある場合は簡単に平均をとり、喫煙率に関してはどうしようもなかったため男性のみのデータを用いた。

男女別のデータについて、本当はデータ収集人数などによる重み付けを行いたかったが、作業量が膨大となるため諦めた。

# データについて④

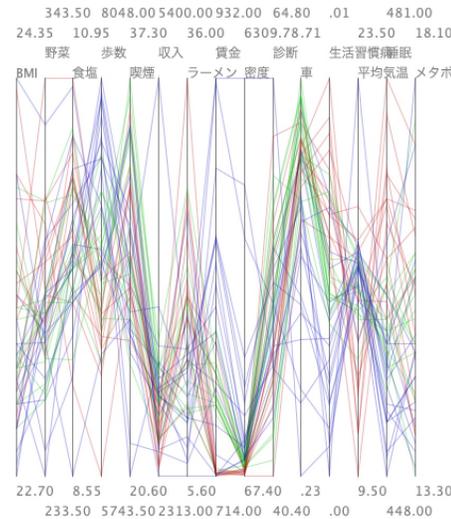
統計データ選定の基準:

- BMI,野菜摂取量、食塩摂取量、歩数、喫煙、健康診断への参加率、生活習慣病患者の割合、メタボの人の割合、睡眠時間は、その都道府県の健康意識を反映しているのではないかと考えた。
- 収入の大小と喫煙の有無は、関係しているのではないかと考えた(この段階ではあくまで偏見である)
- 収入と同様に、最低賃金も喫煙の有無と関係しているのではないかと考えた。
- 一人当たりの車の保有台数と、歩数は関係しているのではないかと考えた。
- 食塩摂取量が多い都道府県はしょっぱいものを好む=一人当たり存在するラーメン店の数が多いのではないかと考えた。
- 気温は、人々の生活にはその土地の気温が大きく関係するのではないかと考えた。
- 県民性に直結するのは田舎か都市部かに関係するのではないかと、そしてこれらは人口密度が表しているのではないかと考え、人口密度をNumericデータにも、Categoryデータにも入れた。

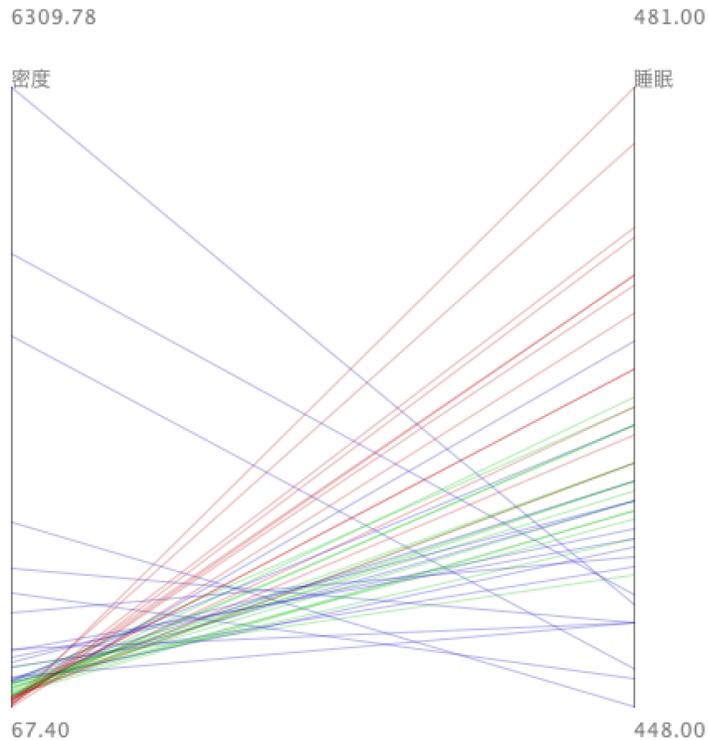
# 結果の外観



まずは外観として散布図を見ていこう。  
予想通り、メタボ、BMI、生活習慣病、歩数、睡眠などには  
相関があるようだ。だがBMIと食塩や野菜摂取量などは離れた位  
置にあるので少々先行きが不安である。  
全体の折れ線図も載せておくが、ここからパッとわかることは  
少なかったため、早速二つのスライダーを動かしていこうと思う。



# 観察①

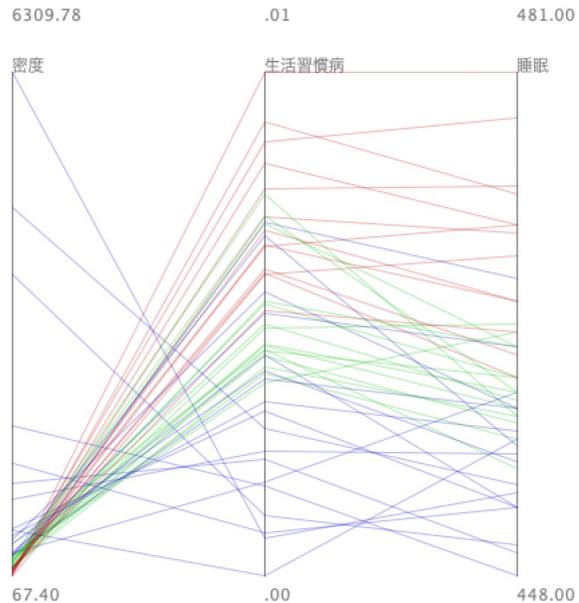


まず右のスライダーを動かしていくと密度と睡眠がつながった。

確かに左の図を見ると負の相関がある。

人口密度が多いということは都市部であるということだから、都市部の方が働く人口が多く睡眠時間が少なくなっているというのは納得できる結果である。

# 観察②

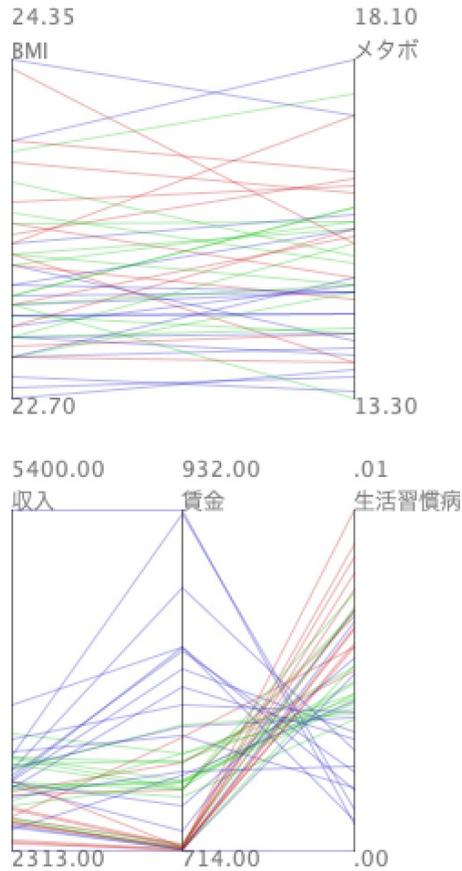


引き続き右のスライダーを上げていくと、次は生活習慣病、密度、睡眠の3つが結ばれた。

ここで驚くことは、生活習慣病と睡眠に正の相関があるということだ。一般的に睡眠時間が少ない方が健康状態が悪化するというイメージがあるが、この正の相関はこのイメージと矛盾するものである。単純に共起したもののなのか、何かしらの因果関係があるのか気になった。

一瞬、都市部よりも地方の方が高齢化が進んでいるた、生活習慣病の患者数が多いのではないかと思ったが、老人の方が生活習慣病に罹りやすいかどうかは分からない。→一般的に生活習慣病にかかりやすい30~50代の人数の統計データを加えてみることを思いつく。

# 観察③



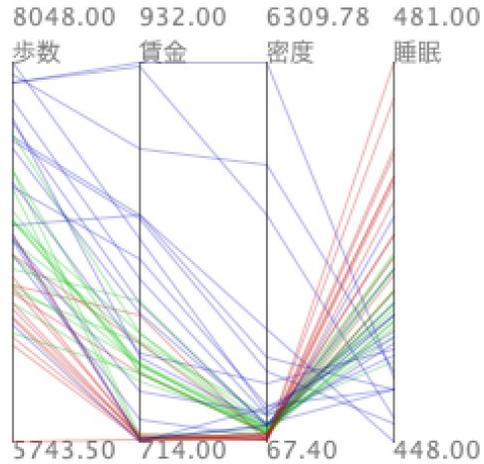
次に結ばれるのは、BMIとメタボ、収入と賃金と生活習慣病である。

まず、BMIとメタボには正の相関があった。BMIに関して女性のデータと男性のデータを雑に足して2で割ってしまったため、ちゃんとしたデータとして扱って大丈夫か不安であったが、[メタボ人口が多い=BMIの平均値が高い]という予想通りの結果が出て安心した。

賃金と生活習慣病について負の相関がある。しかし前のページでも述べたが、生活習慣病が何に起因するものなのか今の段階では明らかではないため、単純に地方で賃金が低く都市部で高いという結果を表しているに過ぎない気がする。

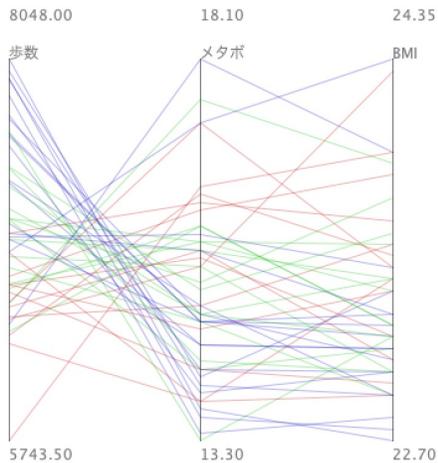
収入と賃金に関しては予想通り正の相関がある。収入と賃金をつなぐ一番上の線は東京である。

# 観察④

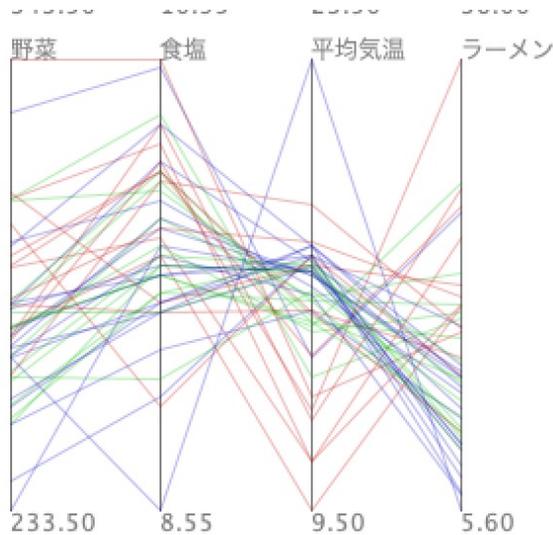


もう少しスライダーを上げると歩数と賃金の正の相関が出てくる。これはつまり、県が栄えていれば栄えているほど歩数が多くなるということである。栄えている県には若者が多かったり、働く人が多かったりと、移動をする人が多いことが関係しているのではないかと思った。

次に、歩数とメタボに負の相関がある。これはすごく直感的に理解しやすい相関である。たくさん歩いていけば、消費カロリーが増えメタボになりにくくなるだろう。



# 観察⑤

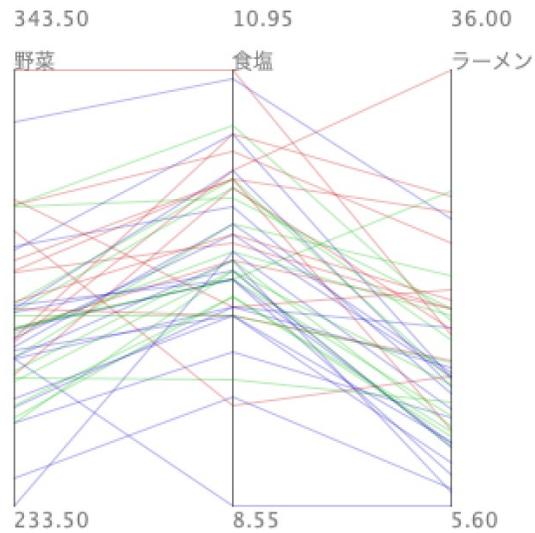


スライダーをさらに上げると、このような折れ線が出てきた。

まず野菜と食塩の関係である。ここにはあまり相関がないように見える。野菜摂取量が多いほど、健康意識が高く食塩の摂取量が少ないかと思ったが、意外にも食塩の摂取量と野菜の摂取量に相関がある。

このグラフで面白いのは平均気温とラーメンの関係だ。都道府県別の健康状態の話からは少し逸れるが、やはり寒い地域の方がラーメンの店舗数が多くなっているようだ。寒いとあったかいラーメンが食べたくなるのを反映しているのかもしれない。この、平均気温が低く、ラーメン店舗は多い人口密度の低い地域は東北地方や北海道かなと思った。→地方ごとにクラス分けしてCategoryデータに入れてみても面白かったかもしれない。(北海道、東北地方は1、のように)

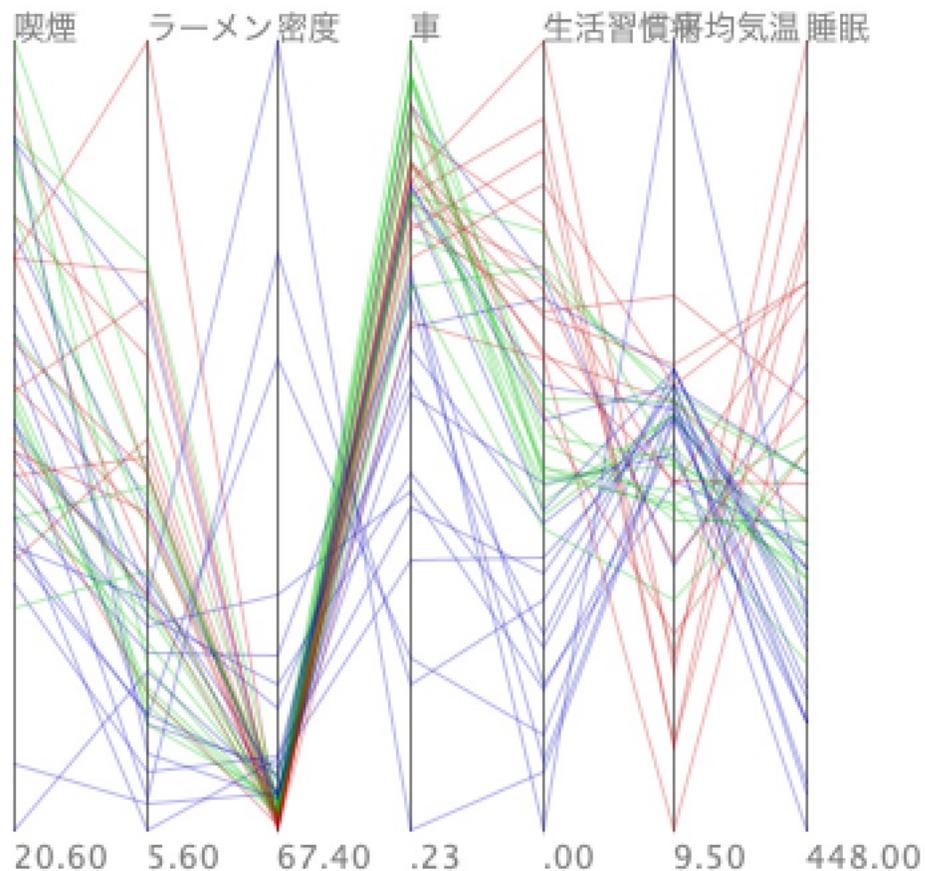
# 観察⑥



ラーメンについての得られた視座をもう一つ。

左の図を見ると、ラーメンと食塩に正の相関がある。やはり、しょっぱいものが好きな県はラーメン好きも多いのだろうか...

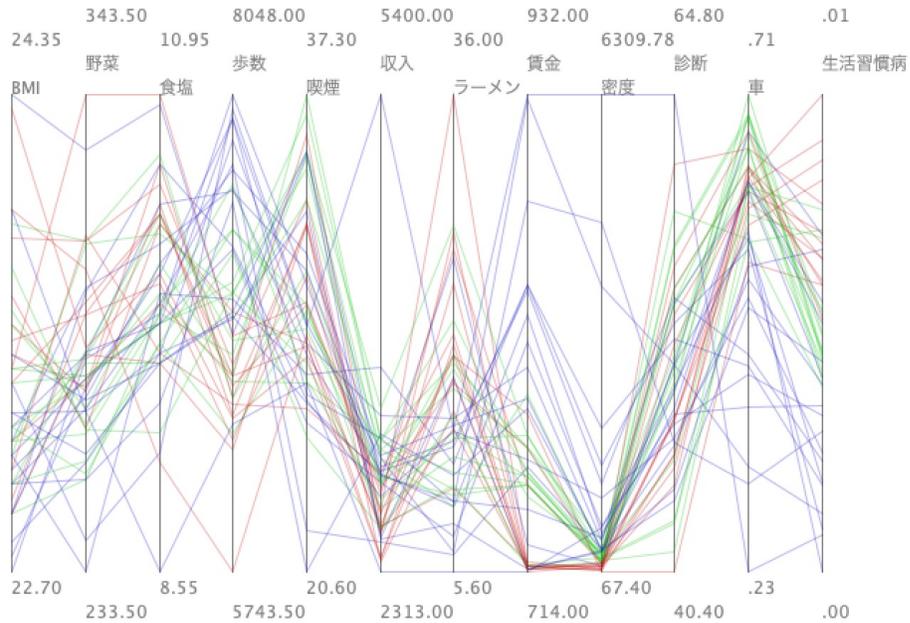
# 観察⑦



さらにスライダーを上げ、車と生活習慣病の関係についてみる。

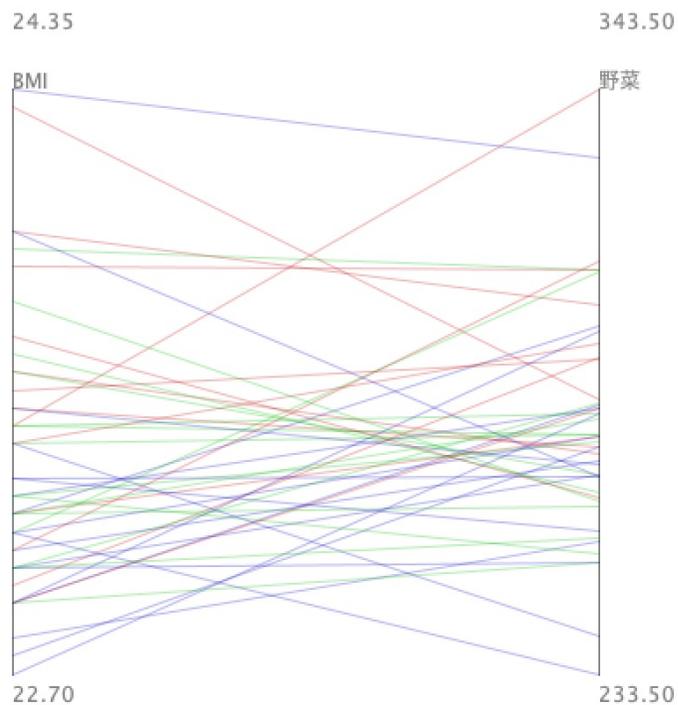
イメージでは車を持っているほど歩くことが少なくなり、生活習慣病にかかりやすいのではないかと思ったが、その結果はこのデータからは得られなかった。

# 観察⑧



スライダーをかなり上げるとここでようやく喫煙のワードが出てくる。この喫煙者数は収入と相関があり、予想では収入が上の方が喫煙者は少くなる=負の相関があると思っていたが、そうではなかった。

# 観察⑨



次に、左のスライダーを動かし、残る特徴量を4つにしてみた。そこからさらにスライダーを動かしてみると、BMIと野菜摂取量の関係が出てきた。

ここには負の相関がある。これは、野菜摂取量が多い県ほど、太っている人が少ないという予想通りの結果を得ることができた。

# 感想①

- 自分の選んだデータが負の相関が出ることを仮定して選んだものばかりなことに後で気づき、かなり相関関係があったのが反省である。
- 思った通りの結果が出ないものが多く時間をかけただけに非常に残念だった。
- データの相関を見たところでそれが因果関係があるのかそれとも単純に共起しただけなのかが分からなかった。